Déconvolution parcimonieuse sans grille: une méthode de faible rang

Paul Catala, Vincent Duval, Gabriel Peyré

▶ **To cite this version:**

Paul Catala, Vincent Duval, Gabriel Peyré. Déconvolution parcimonieuse sans grille: une méthode de faible rang. ORASIS 2017 - Journées francophones des jeunes chercheurs en vision par ordinateur, GREYC, Jun 2017, Colleville-sur-Mer, France. hal-01866617

HAL Id: hal-01866617
https://hal.science/hal-01866617

Submitted on 3 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Low-Rank Approach to Off-the-Grid Sparse Deconvolution
Déconvolution Parcimonieuse sans Grille: une Méthode de Faible Rang

Paul Catala
DMA, ENS
pcatala@dma.ens.fr

Vincent Duval
Mokaplan, INRIA Paris
CEREMADE, Univ. Paris-Dauphine
vincent.duval@inria.fr

Gabriel Peyré
CNRS and DMA, ENS
gabriel.peyre@ens.fr

Résumé

On s'intéresse à la résolution numérique du problème de déconvolution sans grille pour des mesures de Radon discrètes. Une approche courante consiste à introduire des relaxations semidéfinies positives (SDP) du problème variationnel associé, qui correspond ici à un problème de minimisation de variation totale. Cependant, pour des signaux de dimension supérieure à 1, les méthodes usuelles de points intérieurs sont peu efficaces pour résoudre le programme SDP correspondant, la taille de celui-ci étant de l'ordre de f_c^{2d} où f_c désigne la fréquence de coupure du filtre et d la dimension du signal. Nous introduisons en premier lieu une version pénalisée de la formulation SDP, dont les solutions sont de faible rang. Nous proposons ensuite un schéma numérique basé sur l'algorithme de Frank-Wolfe, capable d'exploiter efficacement d'une part cette propriété de faible rang, d'autre part l'aspect convolutif du problème ; notre méthode atteint ainsi un coût de l'ordre de $O(f_c^d \log f_c)$ par itération. Nos simulations sont prometteuses, et montrent que l'algorithme converge en k étapes, k étant le nombre de Diracs dans la solution.

Mots-Clés *déconvolution parcimonieuse ; super-résolution ; relaxations SDP ; Frank-Wolfe*

Abstract

We propose a new solver for the sparse spikes deconvolution problem over the space of Radon measures. A common approach to off-the-grid deconvolution considers semidefinite (SDP) relaxations of the total variation (i.e. the total mass of the measure) minimization problem. The direct resolution of this SDP is however intractable for large scale settings, since the problem size grows as f_c^{2d} where f_c is the cutoff frequency of the filter. Our first contribution introduces a penalized formulation of this semidefinite lifting, which has low-rank solutions. Our second contribution is a conditional gradient (a.k.a. Frank-Wolfe) optimization scheme with non-convex updates. This algorithm leverages both the low-rank and the convolutive structure of the involved variable, resulting in an $O(f_c^d \log f_c)$ complexity per iteration. Our numerical simulations are promising and show that this algorithm converges in exactly k steps, where k is the number of Diracs composing the solution.

Keywords *sparse deconvolution; superresolution; SDP relaxations; conditional gradient*

1 Introduction

1.1 Sparse spikes deconvolution

The super-resolution problem aims at accurately recovering a sparse signal from low-resolution and possibly noisy measurements. This is an important challenge in fields such as medical imaging, microscopy or astronomy, where it may be crucial to overcome the physical limits of the resolving power of sensing devices.
Formally, we want to retrieve a d-dimensional discrete Radon measure $\mu_0 = \sum_{k=1}^r a_j \delta_{x_j}$ ($a_j \in \mathbb{R}$, $x_j \in \mathbb{T}^d$) given the discrete observations

$$y = \Phi\mu_0 + \sigma w \in \mathbb{C}^n,$$

where w is some normalized noise, and Φ is a known filter. In this work we focus on the case where Φ is a low-pass filter of cutoff frequency f_c. In particular, for any measure $\mu \in \mathcal{M}(\mathbb{T}^d)$, we have

$$\Phi\mu = \int_{\mathbb{T}^d} \varphi(x) \mathrm{d}\mu(x), \tag{1}$$

where $\varphi(x) = \left[e^{2i\pi\langle k,\, x\rangle}\right]_{|k|\leqslant f_c}$.

1.2 Beurling LASSO

Although this inverse problem is severely ill-posed, sparse estimates can be found by solving the follow-

ing minimization problem, known as the BLASSO [5]:

$$\mu_\lambda \in \operatorname*{argmin}_{\mu\in\mathcal{M}(\mathbb{T}^d)} \frac{1}{2\lambda}\|y-\Phi\mu\|^2 + |\mu|(\mathbb{T}^d), \qquad (\mathcal{P}_\lambda)$$

where the total variation norm is defined as

$$|\mu|(\mathbb{T}) = \sup\left\{\int_{\mathbb{T}} \eta \mathrm{d}\mu \;;\; \eta\in\mathcal{C}(\mathbb{T}), \quad \|\eta\|_\infty \leqslant 1\right\},$$

and naturally extends the ℓ^1-norm of finite-dimensional vectors to the continuous setting of measures. This grid-free approach offers beneficial mathematical insight on the problem, leading to sharp criteria for stable spikes recovery [2, 3, 9]. However, the infinite-dimensionality of $(\mathcal{P}_\lambda)$ poses a numerical challenge.
The dual formulation plays an important role in the practical resolution of the BLASSO. It aims at the recovery of a trigonometric polynomial $\Phi^* p_\lambda$ where p_λ is the solution of the program

$$p_\lambda = \max_{p\in\mathbb{C}^{n^d}} \left\{\frac{1}{2}\|\frac{y}{\lambda}-p\|^2 \;;\; \|\Phi^* p\|_\infty \leqslant 1\right\}. \qquad (\mathcal{D}_\lambda)$$

In the case where Φ is a low-pass filter, knowing this dual certificate then enables accurate recovery using root-finding techniques [3], since primal-dual optimality relations ensure that the support of μ_λ is contained in the set $\{x \;;\; |\Phi^* p_\lambda(x)| = 1\}$, see Figure 1. In practice, we recover the support by extracting the double roots of $1-|\Phi^* p_\lambda|^2$. Amplitudes are then deduced using the relation $\Phi\mu_\lambda = y - \lambda p_\lambda$ (see Algorithm 2).

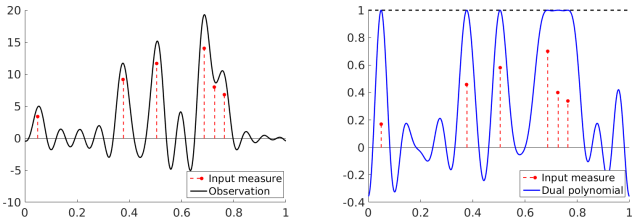

Figure 1 – *Observations (left) and dual polynomial* $\eta_\lambda = \Phi^* p_\lambda$ *(right) for* 5 *Diracs* ($\sigma = 0.002$, $f_c = 13$).

1.3 Previous works

Early works on superresolution [7] considered a discrete approximation of the problem, restraining the search to measures located on a grid. This parameterization enabled the use of basis pursuit methods [4], also known as LASSO.
However, this approximation typically leads to imperfect solutions for thin grids [9], and better results as well as refined mathematical analysis are achieved using the grid-free setting. The support of the measure may then be iteratively estimated using conditional gradient methods [2, 1], or recovered through the dual polynomial, which can be computed in a robust manner using semidefinite relaxations [14, 3, 6].
Although we focus here on ℓ^1-regularization techniques, there is also a vast literature on non-convex superresolution schemes, such as Prony's method [11], MUSIC [13] or FRI-based methods [16], among others.

1.4 Contributions

This work present a novel numerical approach to the superresolution problem. Our method mixes the aforementioned conditional gradient methods with SDP relaxations, thus providing a fast and robust scheme.
Our first contribution introduces a penalized semidefinite formulation of $(\mathcal{P}_\lambda)$ with low-rank solutions that is suitable for conditional gradient schemes. We study the sensitivity of this relaxation.
Our second contribution builds a new numerical solver for this penalized SDP, which fully exploits the aforementioned low-rank structure and the convolutive nature of the problem. In particular, the main operations are performed using Fast Fourier Transforms, thus enabling fast implementation for challenging settings, e.g. superresolution on images.
In the following, although all our results are in dimension 1, the case $d = 2$ currently being implemented, we present the problems in arbitrary dimension.

2 Low-rank SDP relaxation

Although $(\mathcal{P}_\lambda)$ can be used for any degradation operator Φ, we assume in the following sections that it is of the form (1).

2.1 Theoretical background

To deal with the numerical resolution of $(\mathcal{P}_\lambda)$ or $(\mathcal{D}_\lambda)$, it is possible to consider their semidefinite relaxations. SDP hierarchies were introduced by Lasserre [12], and rely on the one hand on a re-parameterization of measures in terms of moment sequences, which remedies the infinite-dimensionality of $(\mathcal{P}_\lambda)$, and on sum-of-squares representations of nonnegative polynomials on the other, which allows to tackle the ℓ^∞-constraint of $(\mathcal{D}_\lambda)$. Although the solutions of these hierarchies have only been proven to converge weakly toward the solution of their original problems, the interest of this method comes from the fact that finite convergence is very frequent in practice. In particular, in many interesting cases, the hierarchy collapses at low order of relaxation.
The Lasserre hierarchy was originally intended for real measures and real polynomials, and its trigonometric equivalent was studied by Dumitrescu [8]. The problems we introduce in the following are derived from his results.

2.2 SDP relaxation

The semidefinite relaxation of $(\mathcal{P}_\lambda)$ at order $(m - n + 1)$ reads:

$$\mathcal{R}_\lambda \in \operatorname*{argmin}_{u,z,\tau} \; h(\mathcal{R}) \overset{\text{def.}}{=} u_0 + \tau + \frac{1}{2}\Big\|\frac{y}{\lambda} + 2z\Big\|^2$$
$$\text{s.t.} \quad \begin{cases} (a) & \mathcal{R} = \begin{bmatrix} R & z \\ z^* & \tau \end{bmatrix} \succeq 0, \\ (b) & R = \sum_{|\boldsymbol{k}|<m} u_{\boldsymbol{k}} \Theta_{\boldsymbol{k}}, \\ & \Theta_{\boldsymbol{k}} \in \mathbb{C}^{m^d \times m^d} \end{cases} \qquad (\mathcal{P}_\lambda^{(m)})$$

where $n = 2f_c + 1$, $m \geqslant n$, $\boldsymbol{k} = (k_1, \ldots, k_d)$, and $\Theta_{\boldsymbol{k}} = \Theta_{k_d} \otimes \ldots \otimes \Theta_{k_1}$. Here Θ_{k_j} denotes the Toeplitz matrix with ones on its k_j-th diagonal and zeros everywhere else, and $\otimes$ stands for the Kronecker product. This formulation follows from the bounded real lemma [8]. The coefficients p_λ of the dual polynomial may then be retrieved from $\mathcal{R}_\lambda$ through the simple optimality relation

$$p_\lambda = \frac{y}{\lambda} + 2z_\lambda. \tag{2}$$

In 1-D, the relaxation is exact with $m = n$, meaning that $(\mathcal{P}_\lambda^{(m)})$ and $(\mathcal{P}_\lambda)$ have the same solution (it results from Caratheodory-Toeplitz theorem [14], or from Fejér-Riesz theorem [8]). In dimension $d > 1$, the relaxation is not tight in the general case, but numerical simulations tend to show that low relaxation orders are sufficient. In particular, as mentioned in [6, Section 4], for $d = 2$, the hierarchy is known to collapse, but with no a priori information on the order m at which it occurs.

As it appears, the size of the above semidefinite program is m^{2d}, with $m \geqslant 2f_c + 1$. Therefore, interior points methods are limited specifically for $d > 1$, or even in 1-D for large values of f_c. The algorithm we detail in Section 4 remedies this issue.

2.3 Low-rank structure

Inspection of the proof in [14, Proposition 2.1] reveals that $(\mathcal{P}_\lambda^{(m)})$ in the univariate case $(d = 1)$ admits a solution whose rank is bounded by the number of spikes in the measure solving $(\mathcal{P}_\lambda)$. Furthermore, this result seems to hold in the d-dimensional case [17]. Our numerical simulations corroborate this low-rank structure, see Fig. 2.

The search space of $(\mathcal{P}_\lambda^{(m)})$ may thus be restricted to matrices of rank at most r, which would allow to store a variable of size $(m^d+1)\times r$ rather than the full matrix $\mathcal{R}$. However, in terms of optimization, the intersection between the manifold of fixed rank matrices and the affine space defined by (b) quickly becomes unanalyzable when $r > 1$ and $n > 2$: for instance, rank-deficient Toeplitz matrices of size $n = 3$ live in the reunion of an affine plane and a cone. In particular, non-convex optimization schemes on this search space seem difficult to implement. Instead, we propose to smooth the geometry of the problem by penalizing the affine constraint (b).

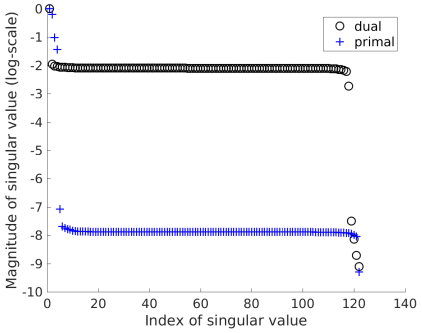

Figure 2 – *($d = 2$, $f_c = 5$ and $r = 4$ spikes). Singular values of the primal (blue) and dual (black) solutions.*

3 Toeplitz penalization

To overcome the difficulty induced by the affine constraint (b), we propose a penalized version of $(\mathcal{P}_\lambda^{(m)})$:

$$\mathcal{R}_{\lambda,\rho} \in \operatorname*{argmin}_{u,z,\tau} h(\mathcal{R}) + \frac{1}{2\rho}\Big\|R - \sum_{|\boldsymbol{k}|<m} u_{\boldsymbol{k}} \Theta_{\boldsymbol{k}}\Big\|^2$$
$$\text{s.t.} \quad \begin{cases} \mathcal{R} = \begin{bmatrix} R & z \\ z^* & \tau \end{bmatrix} \succeq 0, \\ R, \Theta_{\boldsymbol{k}} \in \mathbb{C}^{m^d \times m^d} \end{cases} \qquad (\mathcal{P}_{\lambda,\rho}^{(m)})$$

Note that this approach could equivalently be seen as a perturbation of the atomic norm used in [14] to regularize the problem.

Obviously, the dual polynomial $\eta_{\lambda,\rho}$ deriving from $(\mathcal{P}_{\lambda,\rho}^{(m)})$ (using relation (2) which still holds for the penalized problem) may differ from η_λ, in particular with regard to their roots, thus compromising exact recovery, see Fig. 3. However, following an approach similar to [15], under some mild non-degeneracy hypothesis, it is possible to show that, for small enough values of ρ, $\mathcal{R}_{\lambda,\rho}$ is sufficiently close to $\mathcal{R}_\lambda$ to allow accurate support reconstruction (in particular they have the same rank, see Fig. 4). Numerical observations confirm that this regime exists.

4 Algorithm

In this section, we take advantage of the low-rank property of the solutions as well as of the convolutive structure of the Toeplitz constraint to build a robust and efficient numerical scheme for solving $(\mathcal{P}_{\lambda,\rho}^{(m)})$.

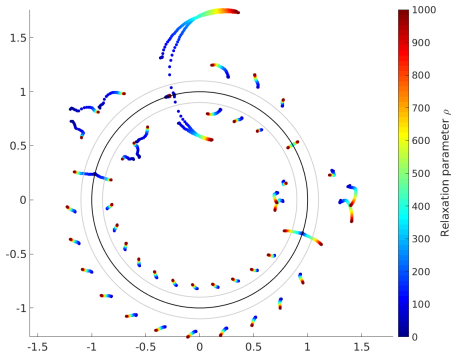

Figure 3 – *($d = 1$, $r = 5$, $\lambda = 0.05$ and $f_c = 13$). Trajectory of the roots of the polynomial $1 - |\eta_{\lambda,\rho}|^2$ w.r.t. ρ. When $\rho = 0$ (constrained case), the support can be recovered from double roots of unit modulus. When $\rho > 0$, these roots may split, but remain identifiable for small values of ρ.*

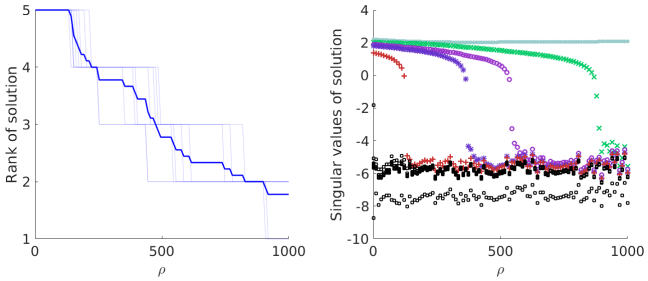

Figure 4 – *($\sigma = 0.005$). Same example as above, with randomized norm. Left: rank of $\mathcal{R}_{\lambda,\rho}$ w.r.t. ρ. We use small perturbations around fixed positions to study the behavior of the rank for similar signals. Right: singular values of one of these signals w.r.t ρ (log-scale). Values below 10^{-4} are taken to be zero.*

In the following, we note

$$f(\mathcal{R}) = u_0 + \tau + \frac{1}{2}\|\frac{y}{\lambda} + 2z\|^2 + \frac{1}{2\rho}\|R - \sum_{|\boldsymbol{k}|<m} u_{\boldsymbol{k}}\Theta_{\boldsymbol{k}}\|^2,$$

where the $u_{\boldsymbol{k}}$ are a function of the coefficients of $\mathcal{R}$.

4.1 Conditional gradient

Conditional gradient, aka Frank-Wolfe (FW) algorithm [10], aims at minimizing a convex and continuously differentiable function f over a compact convex subset K of a Hilbert space. The essence of the method is as follows: linearize f at the current position $s^{(k)}$, solve the auxiliary linear problem of minimizing $s \mapsto \langle \nabla f(s^{(k)}), s\rangle$ on K (Linear Minimization step), and move toward the minimizer to obtain the next position. This scheme ensures the sparsity of its iterates, since the solution after k iterations is a convex combination of at most k atoms.
A remarkable property of this method is that the LM amounts to extracting an extremal point of the set K. When the latter consists in the convex hull of an atomic domain, this subproblem may thus be solved efficiently. In particular, the positive semidefinite cone being generated by rank-one matrices, the LM in this case will result in computing a leading eigenvector of ∇f, which can be done using power iterations.

4.2 Superresolution algorithm

We propose a Frank-Wolfe scheme to minimize the function f over the cone of positive semidefinite matrices. In order to return to optimization over a compact, we set a bound D_0 on the trace of the iterates such that $\operatorname{tr}\mathcal{R}_{\lambda,\rho} \leqslant D_0$. In practice, one may choose $D_0 = f(0)$. To take advantage of the low-rank structure of the solutions, we store our iterates as $\mathcal{R} = ss^H$ (s^H being the conjugate transpose of s). We add a non-convex correction step similar to [1], which consists in a gradient descent on $F : s \mapsto f(ss^H)$. We used a BFGS descent in our implementation, described in Algorithm 1.

Algorithm 1 Recovering dual polynomial

set: $s_0 = [0 \dots 0]^\top$,
D_0 s.t. $\operatorname{tr}(s^\star) \leqslant D_0$ (bound on the domain)
for $r = 1 : N$ (where $N \geqslant 2f_c + 1$ is fixed) **do**
1. Compute next atom:
$v_r = D_0 \arg\min_{\|v\|\leqslant 1} v^\top \cdot \nabla f(s_r s_r^H) \cdot v$
2. Update: $\hat{s}_{r+1} = [\alpha_r s_r, \beta_r v_r]$, with
$\alpha_r, \beta_r = \arg\min_{\alpha+\beta\leqslant 1} f(\alpha s_r s_r^H + \beta v_r v_r^H)$
3. Non-convex corrective step:
$s_{r+1} = \textbf{descent}((s, F(s)) : s \in \mathbb{C}^{(n+1)\times(r+1)})$
end for
return $p_\lambda = \frac{y}{\lambda} + 2z$, where $\begin{bmatrix} z \\ \tau \end{bmatrix}$ is the last column of $s_{N+1}s_{N+1}^H$.

Both the positions and amplitudes of the solution may then be recovered from the coefficients p_λ. Algorithm 2 gives the main steps of the reconstruction process. To recover the roots of $\eta_{\lambda,\rho}$ which correspond to the support of the solution μ_λ (the true candidates), we fix a tolerance around the unit circle in which they should be located, for small enough values of ρ, see Fig. 3.

Algorithm 2 Reconstruction via root-finding

input: p_λ coefficients of the dual polynomial
set: a tolerance $\varepsilon > 0$, R = roots(p_λ)
- **identify candidate roots**:
$R1 \leftarrow R(\big||1 - |R|| < \varepsilon\big)$
- **compute positions (angles of the roots)**:
$\Theta \leftarrow \operatorname{sort}(\arg(R1))$
$x \leftarrow \frac{1}{2\pi}\Theta(1:2:end)$ *# remove symmetric roots*
- **compute amplitudes**:
 # optimality relation
$a \leftarrow (\Phi_x^*\Phi_x)^{-1}\Phi_x^*(y - \lambda p_\lambda)$, where $\Phi_x = \Phi|_x$
return x, a

4.3 FFT-based computations

As mentioned in section 4.1, step 1. in Algorithm 1 simply amounts to computing a leading eigenvector of $\nabla f(s_r s_r^H)$, using power iterations, which is based on matrix-vector products only. Here, given the particular form of f, the sole costly operation will consist in

(i) computing the projection of $s_r s_r^H$ on Toeplitz (or block-Toeplitz in d-dimensional cases) matrices,

(ii) multiply the latter with a vector.

Both steps can be done efficiently using only Fast Fourier Transforms: indeed, the projector in (i) has a simple factorization involving only padding, unpadding, FFT and inverse FFT operators; on the other hand, the Toeplitz-vector product of (ii) may be implemented with the same operators, since it corresponds to a convolution with padding.

4.4 Complexity

As we mainly implemented the 1-D setting for now, we focus here on the complexity for the 1-D case. At step r in algorithm 1, the iterate s_r has size $(n+1)\times r$, with $n = 2f_c + 1$. Step 1 is done with power iteration, using the FFT-scheme described in section 4.3, and costs $O(r\cdot f_c \log f_c + C_{\mathrm{PI}}\cdot f_c \log f_c)$, where C_{PI} depends on the tolerance chosen for power iterations. The line-search in Step 2 has a closed-form expression, hence a constant cost. Finally, the complexity of BFGS iterations in Step 3 basically reduces to the complexity necessary to compute the gradient of $F : s \mapsto f(ss^H)$. Since this can also be done using FFT, the complexity we reach is $O(C_{\mathrm{BFGS}} \cdot f_c \log f_c)$.

5 Numerics

In our simulations, we fixed a limit of 500 BFGS iterations, regardless of the chosen tolerance. The results that are presented in this section were all obtained in 1-D settings.

It appears that the performance of the algorithm does not rely on high BFGS precisions, which is why in practive we often set the tolerance to 10^{-2} or 10^{-3} for this step, see Fig. 5.

Accurate reconstruction essentially relies on the ability to identify the correct roots among all of the roots of $\eta_{\lambda,\rho}$. A possible way would be to use some clustering algorithm to make a distinction between roots that are sufficiently close to the unit-norm and others. In our experiments, we set manually a confidence interval around the unit circle (the value $\varepsilon = 0.1$ in Algorithm 2 worked well in general, but may need to be tuned according to the different parameters), but we will investigate more reliable criteria in future works. Fig. 6 measures the decay of the distance between the

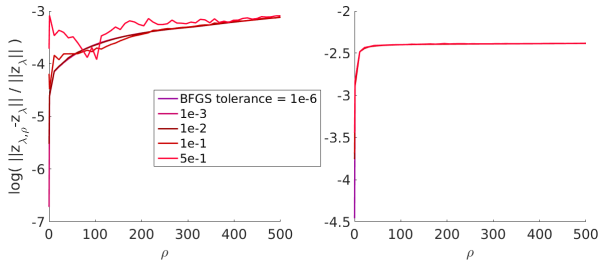

Figure 5 – *($r = 5$, $\lambda = 0.05$, and $f_c = 13$). Precision on dual coefficients for noiseless (left) and noisy ($\sigma = 0.005$, right) observations, in logarithmic scale. In the noiseless example, a tolerance of 10^{-2} appears to be sufficient. In the noisy example, lower precision is reached, but poor resolution on BFGS is sufficient.*

r-th and $(r+1)$-th closest roots to the unit circle as ρ increases for a measure with r spikes.

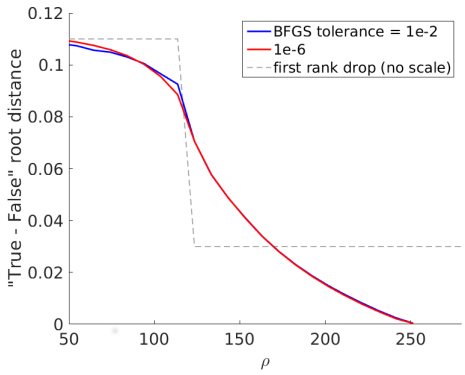

Figure 6 – *($r = 6$, $f_c = 13$, $\sigma = 0$). We measure the distance that separates the worst true candidate from the closest wrong candidate, among the roots of the dual polynomial. This is a fundamental criterion to ensure that we recover the right number of spikes.*

Note that the relation between the roots of the dual polynomial and the rank of the matrix $\mathcal{R}_\lambda$ is unclear; in particular the critical value ρ_c for which we lose tracking of a 'good' root is larger in practice than the value at which the rank drops for the first time.

Finally, an important property of Algorithm 1 is that thanks to the non-convex descent step, it has finite convergence. Even more remarkably, convergence in k steps when k is the number of spikes in the solution is almost always observed with reasonable BFGS tolerances, and for values of ρ that are not too high, see Fig. 7. However, when ρ tends to zero, the method converges more slowly, and in particular the number of BFGS iterations necessary to reach a given tolerance greatly increases in those cases.

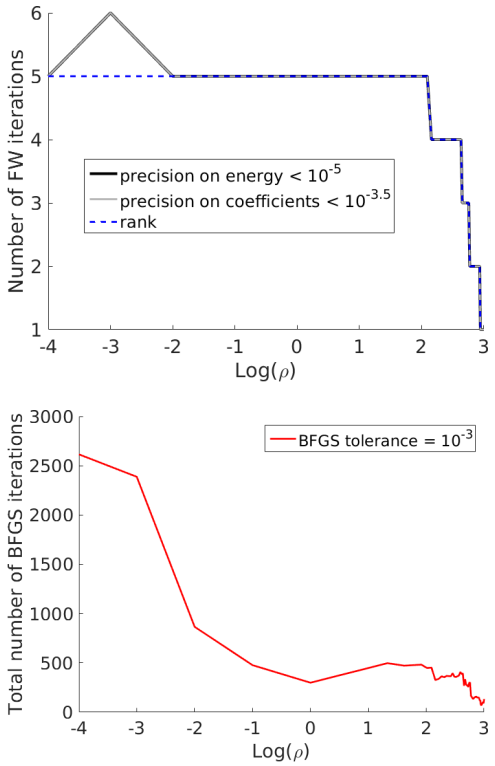

Figure 7 – *($r = 5$, $f_c = 13$ and $\sigma = 0$). Up: Number of Frank-Wolfe iterations w.r.t.* $\log(\rho)$. *This number is computed as the number of iterations necessary to reach a certain precision on either the energy or the coefficients of the dual polynomial. For* $1 \leqslant \rho \leqslant 200$, *we see that the algorithm converged in exactly 5 steps. Bottom: total number of BFGS iterations (summed over all FW iterations) w.r.t.* $\log(\rho)$.

6 Conclusion

The proposed algorithm, being generalizable to 2D, opens new perspectives for the superresolution of images. Owing to its low complexity, our method scales well with the dimension of the problem, and may yield a good alternative to MUSIC or Prony's method in the bidimensional case.

Acknowledgments This work was supported by grants from Région Ile-de-France.

References

[1] N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. In *CAMSAP*, pages 57–60, 2015.

[2] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimization and Calculus of Variations*, 19(1):190–218, 2013.

[3] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001.

[5] Y. de Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical analysis and applications*, 395(1):336–354, 2012.

[6] Y. de Castro, F. Gamboa, D. Henrion, and J.B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Information Theory*, 63(1):621–630, 2017.

[7] D. L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.

[8] B. A. Dumitrescu. *Positive trigonometric Polynomials and Signal Processing Applications*. Signals and Communication Technology. Springer Publishing Company, Incorporated, 2007.

[9] V. Duval and G. Peyré. Sparse regularization on thin grids I: the Lasso. *Inverse Problems*, 33(5), 2017.

[10] M. Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *ICML*, volume 28, 2013.

[11] S. Kunis, T. Peter, T. Römer, and U. Von der Ohe. A multivariate generalization of prony's method. *Linear Algebra and its Applications*, 490:31–47, 2016.

[12] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[13] W. Liao and A. Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *CoRR*, 2014.

[14] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Information Theory*, 59(11):7465–7490, 2013.

[15] S. Vaiter, G. Peyré, and J. Fadili. *Low Complexity Regularization of Linear Inverse Problems*, pages 103–153. Springer International Publishing, 2015.

[16] X. Wei and P.L. Dragotti. FRESH – FRI-based single-image super-resolution algorithm. *IEEE Trans. Image Processing*, 25(8), 2016.

[17] Z. Yang, L. Xie, and P. Stoica. Vandermonde decomposition of multilevel toeplitz matrices with applications to multidimensional super-resolution. *IEEE Trans. Information Theory*, 62(6):3685–3701, 2016.